SCPISM: Review, Implementation and Technical Details

This section is a complete description of the SCPISM as implemented into the CHARMM program, although it contains sufficient details to help implement the model into any other package. Note that each definition and approximation mentioned below is properly justified as discussed in previous publications.

Electrostatic Energy

In the SCPISM the electrostatic energy of a molecule composed of *N* atoms is given by

$$E_{elec} = \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j}{D(r_{ij})r_{ij}} + \frac{1}{2}\sum_{i=1}^{N}\frac{q_i^2}{R_i}\left[\frac{1}{D(R_i)} - 1\right] \qquad (1)$$

where

$$D(r_{ij}) = \frac{1+\varepsilon_w}{1+k\exp(-\alpha_{ij}r_{ij})} - 1$$

$$D(R_i) = \frac{1+\varepsilon_w}{1+k\exp(-\alpha_i R_i)} - 1$$

are the screening functions, with $k = (\varepsilon_s - 1)/2$, ε_s is the static dielectric constant of the bulk solvent, and $\alpha_{ij} = (\alpha_i \alpha_j)^{1/2}$, where α_i is the screening parameter. Formally, these parameters depend on *each atom i* in the molecule (see 'Formal Theory of Polar Media'); however, the current implementation introduces only one parameter per chemical atom type as defined in the CHARMM topology files (*top_*.inp*).

In the SCPISM the effective Born radius R_i of an atom *i* in the macromolecule is given by $R_i = R_{i,w}\xi_{i,w} + R_{i,p}\xi_{i,p} + R_{i,A}\xi_{i,A}$, with $\xi_{i,w} + \xi_{i,p} + \xi_{i,A} = 1$, where $\xi_{i,x}$ is the fraction of atom *i* exposed to phase *x* (*w*=solvent; *p*=solute; *A*=acceptor) and $R_{i,x}$ can be thought of as the Born radius of atom *i* in each phase *x*. For atoms other than polar hydrogens $\xi_{i,A} = 0$ by definition and, then, $R_i = R_{i,w}\xi_{i,w} + R_{i,p}\xi_{i,p}$, with $\xi_{i,w} + \xi_{i,p} = 1$.

$R_{i,w}$ is defined by $R_{i,w} = R_{i,COV} + \delta_i(q_i)$, where $R_{i,COV}$ is the covalent radius of atom *i* and $\delta_i(q_i)$ is an extension that depends, in general, on each atom in the molecule and its partial charge q_i. In the current implementation the quantity $\delta_i(q_i)$ depends only on the sign of the charge, not on its magnitude, and is independent on the atom, i.e., $\delta_i(q_i) \equiv \delta(\text{sign}(q_i))$, and given by $\delta(+) = 0.35\,\text{Å}$ and $\delta(-) = 0.85\,\text{Å}$. In addition, only five values of $R_{i,COV}$ are introduced, one for each chemical element C, O, N, S, and H.

In analogy with $R_{i,w}$, the radius $R_{i,p}$ is defined as $R_{i,p} = R_{i,COV} + \gamma_i(q_i)$, where γ is the extension of the atom in its particular molecular environment. In the current implementation γ is approximated as $\gamma_i(q_i) = \delta(sign(q_i)) + \lambda$, where $\lambda = 0.5\,\text{Å}$ for *all* atoms in the molecule.

Note: all the approximations mentioned above might require revision in the future.

The quantity $R_{i,A}$ is as essential for the quality of the results as the screening parameters α_i that controls bulk electrostatics (as defined in $D(x)$ above). These coefficients allow for fine-tuning the hydrogen bonding interactions in the system. Unlike $R_{i,w}$ and $R_{i,p}$ that are *calculated* based on the definitions above, $R_{i,A}$ are empirically *adjusted* to target the right hydrogen-bonding (HB) energy between a donor and acceptor group that share the proton (polar hydrogen) *i*. Note also that these quantities are expressed as $R_{i,A} = (g_{i,a} g_{i,d})^{1/2}$ where $g_{i,x}$ is a coefficient associated to the acceptor ($x=a$) and to the donor ($x=d$) of the shared proton *i*. Thus, the correct HB energy depends on the donor and acceptor groups as detailed in previous publications. In the general case $g_{i,x}$ are classified according to the chemical nature and hybridization sates of the donor and acceptor atoms, and the net charge of the acceptor and donor groups. Moreover, in the original implementation for use in MC calculation an angular dependence was also introduced that account for the HB geometry. For MD simulations (the current implementation into CHARMM) this angular dependence was removed. The current implementation also simplifies the

number of $g_{i,x}$ assigned to the acceptor and donor atoms, from 42(*s*-*s*)+13(*s*-*b*)+1(*b*-*b*)=56 different HB classes, to only 10(*s*-*s*)+1(*b*-*b*) classes (*s*-*s* denotes side-chain/side-chain interactions; *s*-*b*, side-chain/backbone; and *b*-*b*, backbone/backbone).

Although the fractions $\xi_{i,x}$ can be calculated numerically, as done in [] for MC calculations, the current version of the SCPISM for MD simulations uses a contact-like model approximation, i.e.,

$$\xi_{i,w} = \left[A_i - \sum_{j \neq i}^{N} B_{ij} \exp(-C_{ij} r_{ij}) \right] \Big/ 4\pi (R_p + R_{i,vdW})^2 \qquad (2)$$

where R_p and $R_{i,vdW}$ are the probe radius (equal to 1.4Å) and the van der Waals radius of atom *i*, respectively. In the current implementation a simplification is made where the coefficients B_{ij} and C_{ij} depend only on the central atom *i*, i.e., $B_{ij} \equiv B_i$ and $C_{ij} \equiv C_i$. For non-polar hydrogen atoms the fraction $\xi_{i,p}$ is obtained from $\xi_{i,p} = 1 - \xi_{i,w}$. For a polar hydrogen the fraction exposed to the acceptor atoms, $\xi_{i,A}$, is subtracted from the fraction exposed to the solute, $\xi_{i,p}$, i.e.,

$$\xi_{i,p} = 1 - \xi_{i,w} - \sum_{j \neq i}^{M} D_{ij} \exp(-E_{ij} r_{ij}) \qquad (3)$$

$$\xi_{i,A} = \sum_{j \neq i}^{M} D_{ij} \exp(-E_{ij} r_{ij}) \qquad (4)$$

and $\xi_{i,w}$ is still defined by Eq.(2). In Eqs.(3) and (4) the sums run over all the acceptor atoms (*M* is the total number of acceptors). In the current implementation the coefficients depend only on the central atom *i*, i.e., $D_{ij} \equiv D_i$ and $E_{ij} \equiv E_i$. In practice, the summations in Eqs.(2)-(4) are restricted to atoms within a cutoff distance r_{cutoff} of the central atom *i*, which in the current implementation is set to $r_{cutoff} = 6$ Å.

Coefficients A_i, B_i, C_i, D_i and E_i in Eqs.(2)-(4) are optimized as described in previous publications. Note that these coefficients are not part of the SCPISM *per se* but a complement to it, that account for a particular aspect of the topology of the molecule.

Values of parameters α_i (electrostatics) and $R_{i,A}$ (hydrogen bonding) depend strongly on the force field used (for the current implementation in CHARMM version c31b1 see *scpism.inp* parameter file). The coefficients A_i-E_i also depend on the force field used because they contain information on the $SASA_i$ and, then, on the van der Waals radii of the atoms.

Solvation Energy

The polar component of the solvation energy is given by

$$\Delta G_{elec} = \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j}{r_{ij}}\left[\frac{1}{D(r_{ij})}-\frac{1}{D_0(r_{ij})}\right]+ \\ +\frac{1}{2}\sum_{i=1}^{N}q_i^2\left\{\frac{1}{R_i}\left[\frac{1}{D(R_i)}-1\right]-\frac{1}{R_{i,0}}\left[\frac{1}{D_0(R_{i,0})}-1\right]\right\} \quad (5)$$

where $D_0(x)$ and $R_{i,0}$ are the analogous to the screening $D(x)$ and Born radius R_i but when the molecule is in the vacuum. Although a position-dependent form is also expected for $D_0(x)$, in the SCPISM these screening functions were assumed to be constant throughout the system, i.e., $D_0(x) = D_I$. In this case D_I can be approximated by the static dielectric constant ε_I of the molecule in the gas phase, which is determined mainly by the electron structure. This approximation is reasonable for small systems such as those used in the parameterization of the model, i.e., amino-acids side-chains analogs. Values for ε_I can be obtained from *ab initio* methods or from available experimental data. However, for the parameter optimization in CHARMM they were treated as open parameters, on the same foot as α_i.

Parameters Optimization:

For the parameterization of the SCPISM twenty-one amino-acids side-chains analogs were used. With the approximation $D_0(x) = D_I$, the polar contribution, ΔG^I_{elec}, to the total free energy, $\Delta G^I = \Delta G^I_{elec} + \Delta G^I_{np}$, of each analog *I*, is expressed as

$$\Delta G^I_{elec} = E^I_{elec} - \frac{1}{2D_I}\sum_{i\neq j}^{N}\frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left[\frac{1}{D_I} - 1\right]\sum_{i=1}^{N}\frac{q_i^2}{R_{i,0}} \qquad (6)$$

where E^I_{elec} is given by Eq.(1) for the molecule *I*, and ΔG^I_{np} is the non-polar term. The parameter optimization was based on a simulated annealing Monte Carlo in the space of the parameters $\{\alpha_i, D_I\}$. A function $\mathcal{G}$ is defined by

$$\mathcal{G}(\{\alpha_i\},\{D_I\}) = \sum_I \left[\Delta G^I_{elec}(\{\alpha_i\},\{D_I\}) - \Delta G^I_{elec,0}\right]^2 \qquad (7)$$

where $\Delta G^I_{elec,0}$ is the polar contribution to the *experimental* hydration energy of each molecule. This quantity was taken as $\Delta G^I_{elec,0} = \Delta G^I_0 - \Delta G^I_{np}$, where ΔG^I_0 is the experimental hydration energy and $\Delta G^I_{np} \approx a + b\,SASA_I$ (*a* and *b* are parameters obtained in a prior calculation by fitting experimental hydration energies of the eight alkanes C_nH_{2n+2}). A standard Metropolis criterion in the canonical ensemble with a Boltzmann factor $f = \exp(-\mathcal{G}/T)$ was used; *T* is an arbitrary parameter that decreases according to a standard logarithmic schedule. Stochastic sampling in the space $\{\alpha_i, D_I\}$ leads to a minimization of $\mathcal{G}$; ideally $\mathcal{G} \to \mathcal{G}_0 = 0$ as $T \to 0$ (in practice $\mathcal{G}_0 > 0$).